REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
12 Spe 2012

2. REPORT TYPE
Final Technical

3. DATES COVERED *(From - To)*
1 Aug 2008 - 30 Nov 2011

4. TITLE AND SUBTITLE
Spatial hearing, attention and informational masking in speech identification

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-08-1-0424

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Gerald Kidd

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Boston University

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR
875 N. Randolph St
Arlington, VA 22203
Dr. Willard Larkin/RSL

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-OSR-VA-TR-2012-1081

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution A: Approved for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

During the award period covered by this report (12-1-2008 through 11-30-2011) the research groups at Boston University and at Wright-Patterson Air Force Base have collaborated on portions of the work described in the following progress report. These collaborative efforts have taken several forms including consultation regarding experimental design, jointly conducted experiments, sharing of results, discussions of the interpretation and theoretical implications of research findings, and the planning of new studies. A brief overview of our collaborative projects is given in the following paragraph with specific examples provided throughout the report.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Gerald Kidd
unclassified	unclassified	unclassified	unclassified (U)	24	19b. TELEPHONE NUMBER *(Include area code)* PI phone number

FINAL REPORT: "Spatial hearing, attention and informational masking in speech identification"

I. Collaboration

During the award period covered by this report (12-1-2008 through 11-30-2011) the research groups at Boston University and at Wright-Patterson Air Force Base have collaborated on portions of the work described in the following progress report. These collaborative efforts have taken several forms including consultation regarding experimental design, jointly conducted experiments, sharing of results, discussions of the interpretation and theoretical implications of research findings, and the planning of new studies. A brief overview of our collaborative projects is given in the following paragraph with specific examples provided throughout the report.

The BU and WPAFB groups routinely have held joint discussions of research at the Spring meeting of the Acoustical Society of America, the Midwinter Research Meeting of the Association for Research in Otolaryngology, and at the annual Binaural Bash conference held at Boston University. Furthermore, there have been specific visits for the purpose of fostering collaborative research projects that have occurred outside of these regular scientific group meetings. For example, Dr. Virginia Best was sponsored by a Window on Science Program through the Asian Office of Aerospace Research and Development to visit the group at WPAFB during August 18-29, 2008, for discussions of ongoing collaborative research. An earlier visit and group meetings at the larger scientific society conferences led to a collaboration described in a scientific paper presented at the 2009 meeting of ARO. Work on that project, and related studies, continues.

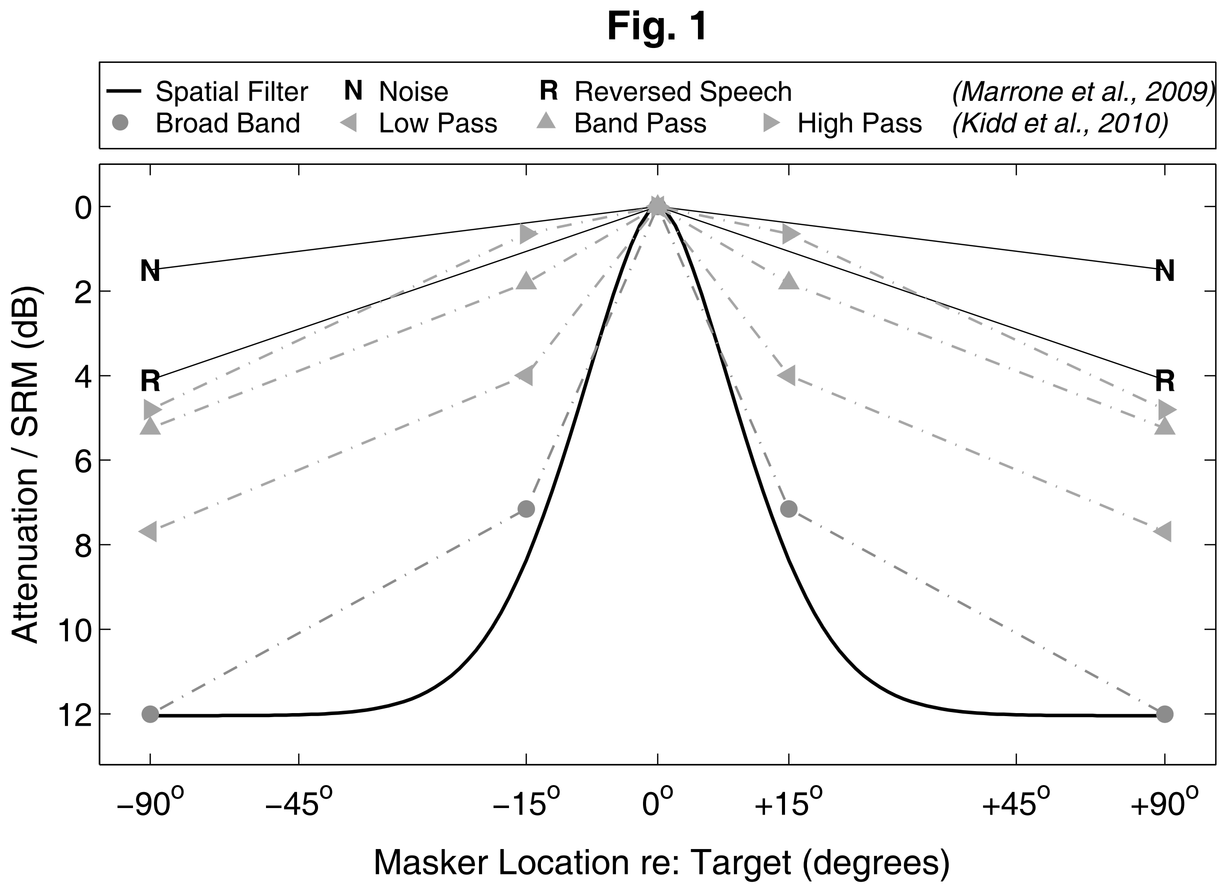
II. Progress towards Specific Aims

II.A. Tuning in the Spatial Dimension

Largely as a consequence of the work supported by this grant during the previous award period, our understanding of "spatial tuning" in azimuth has advanced considerably. The following is a summary of the progress made toward this aim referring where appropriate to both published and unpublished work and also discussing areas that remain unclear or that require further study.

Part of the impetus for formulating this aim was the Ph.D. dissertation of Nicole Marrone (2007) indicating conditions under which highly selective spatial tuning could be observed. A

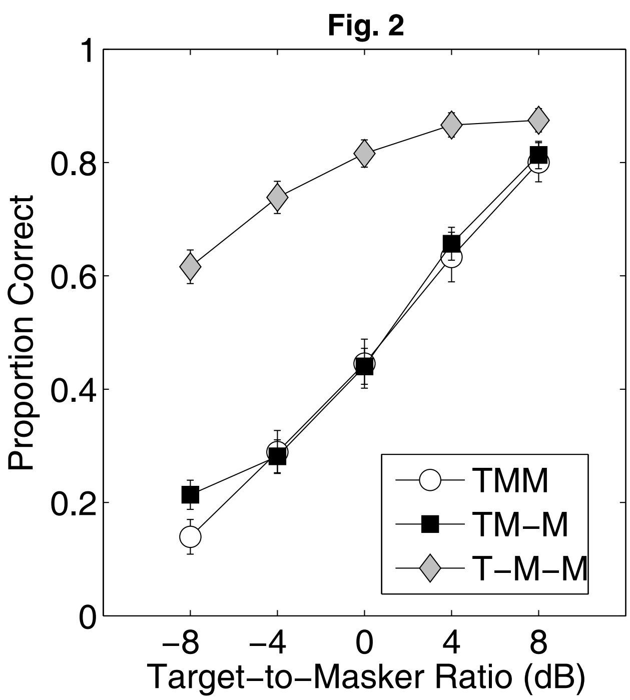
portion of that work and some extensions were subsequently published by Marrone et al. (2008a,b,c). Essentially, the gist of their findings was that highly "tuned" behavioral responses to sound sources located in different spatial configurations could be observed under conditions in which a high degree of informational masking was present. The conclusion that spatial tuning was related to overcoming informational, rather than energetic, masking was the key finding. The reason that this conclusion is key is that it implicated selective attention as the basis for spatial tuning rather than traditional binaural analysis mechanisms (also related see Best et al., 2005; Allen et al., 2008). The term "binaural analysis" is often used as a catch-all for any binaural advantage that is not a consequence of simple acoustics (i.e., differential attenuation of sounds at the two ears due to "head shadow"). Binaural analysis is most often invoked as an explanation for the "masking level difference" (MLD) which provides a robust advantage for the detection of a low-frequency tone in Gaussian noise due to interaural differences between the tone and noise (for topical reviews see Durlach and Colburn, 1978; Colburn, 1995; Stern and Trahiotis, 1996). The parallel reduction in target-to-masker ratio at threshold for speech reception due to similar manipulations in interaural signal and noise parameters may also be related to the same within-channel mechanisms (cf. Levitt and Rabiner, 1967; Zurek, 1993; Culling and Colburn, 2000; Culling et al., 2006). Zurek (1993), for example, calculates that binaural analysis contributes a maximum of about 3-5 dB to the overall spatial release from masking (SRM: the difference between target-to-masker ratio at threshold, T/M, in colocated and separated source conditions) for speech in noise in an anechoic sound field. In realistic sound fields, however, in which some reverberation is present the release is less. Marrone et al. (2008a) have also found a small SRM (about 1.5 dB for symmetrically-separated maskers) in a sound field when the maskers were speech-shaped noise that produced primarily energetic masking. Because that value reflects the maximum attenuation of the spatial filter, the effect of tuning - whether sharp or broad - is minimal. However, when the dominant form of masking is informational masking, much larger SRM is often reported ranging from about 12-18 dB based on several findings from our laboratory (e.g., Arbogast et al., 2002; Marrone et al., 2008a; Kidd et al., 2010). The bandwidth of the spatial filter also may be quite narrow, in the range of 10-15°. Further studies completed recently have helped to clarify this process. Kidd et al. (2010; see also Best et al., 2011) measured SRM for combinations/ proportions of energetic and informational maskers at different spatial separations. They also examined performance under various filtered conditions designed to limit the availability of interaural time and level cues (ITDs and ILDs, respectively). A composite figure illustrating some of these findings is shown in **Figure 1**.



This figure contains a schematic of a hypothetical spatial filter (solid line with no data points) computed using the SRM values from Marrone et al. (2008a) for a three-talker mixture with the target talker always directly in front of the listener (0° azimuth) and the maskers either colocated with the target or symmetrically spatially separated. In addition, two other sets of data are also plotted for comparison: first, the results for maskers located only at 0° or ± 90° for reversed speech and speech-shaped speech-modulated noise are plotted as R and N, respectively. Both the noise and reversed speech are intended as controls for the energetic masking produced by forward speech; however, they are thought to produce less informational masking with the noise producing very little informational masking and the reversed speech an intermediate amount. These data were part of the Marrone et al (2008a) study and were obtained using the same procedures and subjects. The other data points are from subsequent work by Kidd et al. (2010) using similar procedures and stimuli. The parameter that is varied in that study is the filtering of the speech targets and maskers (see symbol key), which were broadband (partial replication of Marrone et al. 2008a), low-passed at 1.5 kHZ, band-passed (1.5-3 kHz) and high-passed (3 kHz).

Overall this figure is intended to illustrate several points regarding spatial tuning: first, highly tuned responses may be observed under conditions dominated by informational masking; this is apparent from the solid filter function based on Marrone et al.'s data and other related work from our laboratory (see also the recent work by Wan et al., 2010, modeling the Marrone et al. data using a modified E-C model). Generally, the magnitude of SRM varies with the amount of informational masking as indicated by the results from the noise and reversed speech maskers. This suggests that large SRM for speech identification is primarily a consequence of release from informational masking. However, in order to achieve this large SRM and sharply tuned responses the observer also must have robust binaural information to use in focusing attention in azimuth. This conclusion is based on the filtered speech results, which were obtained under high informational masking conditions. In those conditions, low-pass filtering limited the usefulness of ILDs and high-pass filtering limited the usefulness of ITDs. Under those filtered conditions, SRM was reduced relative to the broadband case and the pattern of attenuation, to the extent that it could be ascertained, was less sharply tuned. Thus it appears that when binaural information is degraded by limiting useable interaural time or level differences, both the apparent sharpness of tuning and the magnitude of SRM were both significantly reduced.

The preceding work led to questions regarding the extent to which sound sources falling within the focus of attention could be separated perceptually when additional spatially separated sources (presumably outside of the primary focus of attention) were present. The work described in this section developed in stages. Initially, Brungart et al. (2007) reported findings from conditions in which two speech maskers were varied in location relative to a target speech source. In one subset of conditions, which was implemented by presenting stimuli through earphones and applying HRTFs to create spatialized images, one masker was colocated with the target and a second masker was spatially separated. Separating the second speech masker did not improve speech recognition performance relative to the case in which both maskers were colocated with the target. This finding suggested that spatial filtering is compromised when a complex segregation task must be performed at the point of focus of attention; i.e., at the target location. That result, however, appeared to be inconsistent with some unpublished work from the Psychoacoustics Laboratory at BU. In order to understand the reasons underlying the different findings, Dr. Virginia Best visited the laboratory at WPAFB and collaborated on a series of experiments conducted over several days that investigated the role of some of the differences in design between the two studies. A summary of a portion of those results is shown in **Figure 2** (Best et al., unpublished).



In this figure, group mean proportion correct speech identification performance is plotted for three target and masker configurations (always two independent speech maskers): target and maskers colocated (TMM), one masker colocated with the target and one separated (TM-M), and both maskers separated (T-M-M) as a function of target to masker ratio. In general, no advantage was found for separating only a single masker while very large advantages (greater than 10 dB) were apparent when both maskers were separated. These results generally supported the earlier report of Brungart et al. (2007). However, the recently published findings of Kidd et al. (2010) suggest a more complex picture in which a variety of factors influence tuning in mixed-location conditions and suggest that under some conditions significant SRM may occur even when one masker is colocated with the target. In both the Best et al. study and the Kidd et al. study threshold (roughly 50% correct) occurred around 1-4 dB. In the Best et al. study moving one masker away from the target did not improve performance. In the Kidd et al. study, however, 50% correct performance was achieved under a similar combined colocated-separated masking condition for a target-to-masker ratio around -11 dB, yielding a SRM of 12 dB. One crucial variable that appears to underlie the large benefit of moving one masker off the point of focus is the very low threshold found when there was only a single masker talker colocated with the target. The group mean threshold found by Kidd et al. for a 1-talker masker colocated with the target was about -22 dB. Thus, adding a second independent colocated masker talker raised target threshold by 23 dB - an enormous increase in masking. Although the Best et al. study replicated the earlier report by Brungart et al., it did not measure the single masker talker condition when target and masker were colocated. This leaves open the possibility, but does not prove, that the degree of difficulty in the segregation task at the point of focus of attention is key. In the Kidd et al. study, as with a number of related findings, performance when there are two independent speech maskers colocated with a target talker is relatively stable (small variability across subjects and studies) at a slightly positive target-to-

masker ratio. We believe that this positive value is due to a high degree of informational masking and limited cues for segregating the target; for example, reversing the two speech maskers, which preserved the energetic masking but greatly reduced the informational masking present, decreased thresholds by 12 dB consistent with the presence of a high degree of informational masking. However, when only a single masker talker is present, the variability across subjects and studies is quite large and appears to depend heavily on the specifics of the speech materials and presentation conditions. Thus, under some conditions the listener may easily segregate the target and (single) masker in the colocated case leading to the very low threshold T/Ms, such as those found by Kidd et al. In that case, spatial filtering may provide a very large benefit by attenuating the separated talker when a second masker talker is added. If the segregation task is difficult even for one talker colocated with the target (which we speculate may have been the case for the conditions tested by Brungart et al. and Best et al.) then attenuating the second spatially separated talker would have little effect. This issue reveals the complexity of the interactions that may take place among multiple sources and how the magnitude of the advantage of spatial separation depends on the cues available to the listener and the degree of informational masking present.

II.B. Stream Formation, Segregation and Maintenance over Time

This aim is broad encompassing many aspects of current research in hearing. Understanding what comprises an auditory stream and how the listener maintains the linkage between the elements of the stream under the pressure of competing maskers is fundamental to understanding multisource listening. The work in this area attempted to examine which factors bind sounds together to form auditory streams focusing primarily on speech with some of the more recent experiments extended to include nonspeech patterns.

The first work under this aim used a new adaptation of the procedure originally developed by Broadbent (1952). In this procedure, two sentences are presented to the listener in alternating word format. So the words from talker A are the odd-numbered words in the sequence while the words from talker B are the even-numbered words in the sentence. Because the words from the two talkers do not overlap in time there is no simultaneous (or energetic) masking. Our control conditions also indicate that for the parameters used in the experiments any forward masking is inconsequential. Thus, the masking that occurs - which may be quite significant - is all informational masking. This paradigm is very useful for examining speech features or stimulus variables that link sounds together perceptually or

semantically. The initial work using this procedure and a specifically designed speech corpus was reported by Kidd et al. (2008b). One subset of those findings will be reviewed here in detail.

Fig. 3

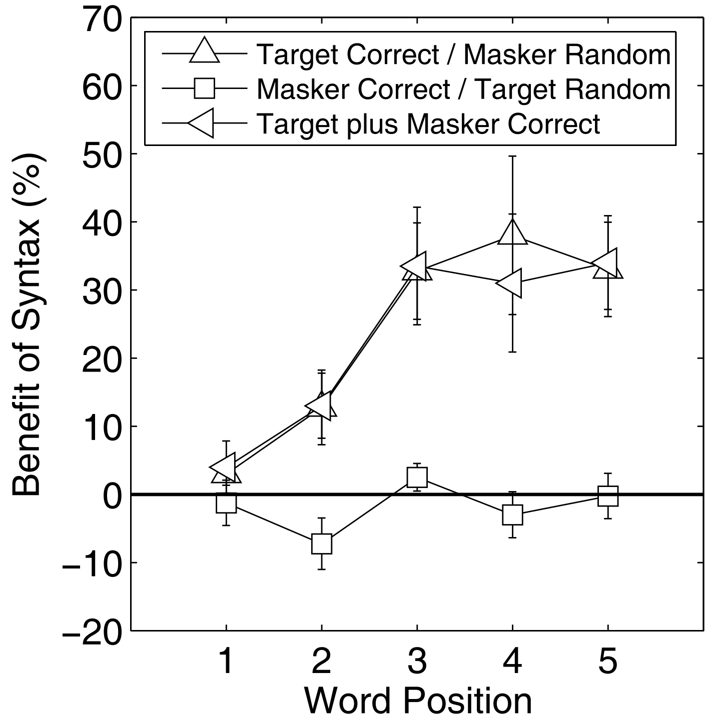
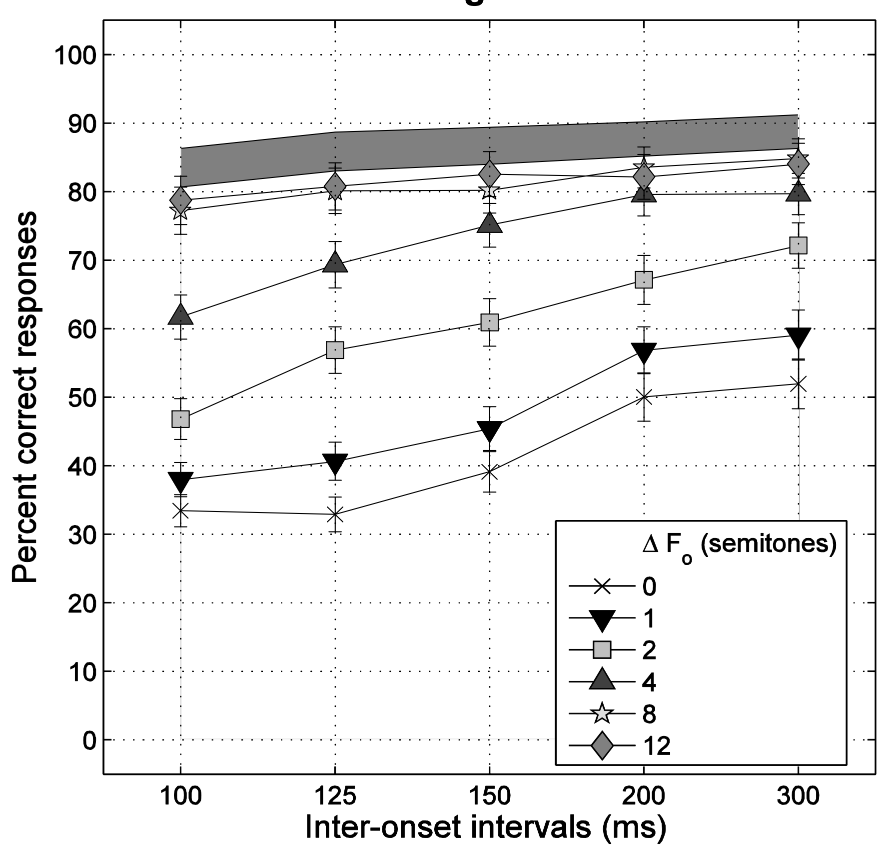


Figure 3 illustrates the benefit in speech identification performance that may be obtained by linking together the target words in a sequence according to correct syntactic structure. This figure shows the advantage in percent correct word identification due to imposing correct syntax (relative to words chosen randomly without regard to syntax) as a function of the position within the 5-word sentences. The legend indicates whether the target sequence, masker sequence or both were syntactically correct and the benefit is relative to the case in which both sequences have random word order. First, large benefits to performance were found when the target sentence was syntactically correct. Thus, expectation about the target word order provided a large advantage over random order with the greatest benefit occurring for words 3-5 within the sentence. Second, no benefit was found when the masker words were ordered correctly for a random target. This finding is consistent with other manipulations tested in the study which varied masker predictability; only manipulations linking the target words together were beneficial. This latter point is considered further below in the context of the Listener Max-Min observer strategies. The finding regarding the benefit of correct syntax in overcoming informational masking is important because it reveals that observer expectation may play a critical role in stream maintenance and the reception of information conveyed by the stream. Target words presented in random order test serial recall. Imposing syntactic order to word sequences increases the predictability of the words and conforms to the rules of normal language. While it is expected that recall would be better for syntactically presented words than randomly ordered words, the significant finding here is the difference in the vulnerability of the materials to informational masking. This finding suggests that listener expectation may be very important in multisource listening under high informational conditions.

A second project examining streaming was conducted jointly with the WPAFB group. Iyer et al. (2009) used the speech corpus developed by Kidd et al. (2008b) in a novel experimental paradigm in which listeners identified target words played in repeating loops. As in the adaptation of the Broadbent (1952) procedure described above, masker stimuli were temporally interleaved with the target words. The maskers varied in informational masking content including noise, reversed speech and forward (intelligible) speech. In addition to the informational masking value of the maskers, stimulus parameters promoting grouping/segregation were manipulated. These included fundamental frequency differences between speech targets and maskers (in semitones) and inter-onset interval for the 100-ms duration test and masker items. These parameters assessed performance under conditions similar to those used in the common A-B-A tone sequence streaming experiments (cf. van Noorden, 1975; review in Bregman, 1990). **Figure 4** shows some of the findings from this study. The ordinate is group mean percent correct identification of the target words while the abscissa is time between item onsets. The shaded region at the top of the graph shows identification performance when the masker was reversed speech having the same fundamental frequency as the target (0 semitones F_0 difference), which forms a reference for comparison with the high-informational masking forward-masker speech results. Noise maskers had little effect on intelligibility and those results are not shown. The functions indicate performance for

Fig. 4



differences in fundamental frequency between target words and masker words. This result is in agreement with the Kidd et al. (2008b) findings regarding the benefit of a constant target voice in reducing the informational masking produced by competing voices. As seen in Figure 4, the greater the separation in F_0 the better the performance at any temporal separation with improving performance found as the time between elements was lengthened. The largest advantage of forward- vs. reversed-masker speech is indicated by the difference between the shaded region at the top and the lowest function which were both measured at 0 semitones F_0 separation. These findings demonstrate that stream segregation for sentences behaves in a similar manner to the A-B-A tone-sequence streaming results (e.g., Bregman, 1990) in that performance improves with increasing frequency/F_0 separation; however, they differ in that listeners were better able to process the target speech stream when rate was slower possibly due to linguistic processing factors.

II.C. Develop and a Test Quantitative Model of IM: Ideal Processing, Acceptance vs. Rejection Processing and Stimulus Uncertainty

The work in this section continued a line of research into the role of two hypothetical observer strategies that describe how attention may act to improve source selection. The theory behind this work is an extension of the conceptual framework originally developed by Durlach (1963; 1972) and embodied in the equalization-cancelation model (EC). The basic idea, as applied to the role of attention in source selection, is relatively simple: the observer controls filters distributed along a particular stimulus dimension(s); for example, frequency or azimuth. In one mode of operation, the observer selects the filter containing the signal and enhances the output of that filter relative to other filters. This represents an acceptance-filter approach also called Listener Max (L_{Max}) because it maximizes performance within the desired filter(s). In contrast, the observer can apply rejection filtering or "nulls" at locations where undesired sources are present, reducing their outputs. This approach is termed Listener Min (L_{Min}) because the processing minimizes the effect of unwanted sources. Both approaches may be useful in segregating/selecting a target source among competing maskers. Although a simple filter selection process is the example given here acceptance or rejection filtering may be much more complex and the conceptual framework has proved to be useful in a wide variety of applications, many of which involve spatial processing of sound sources (e.g., Akeroyd, 2004; Gallun et al., 2005).

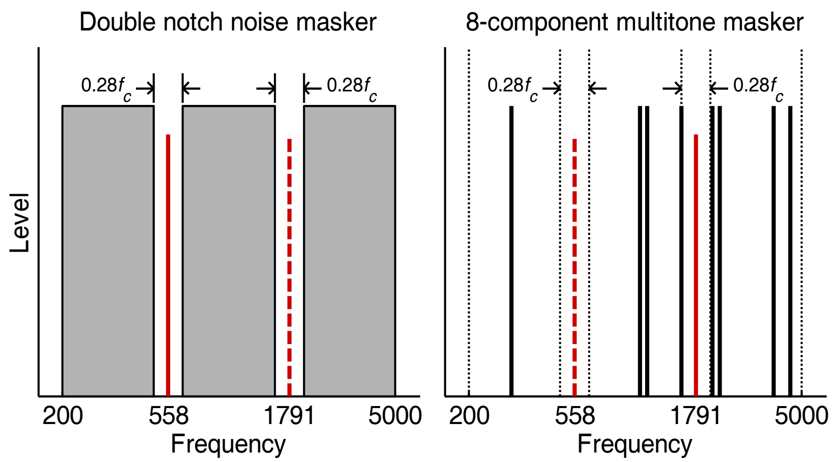
During the past award period, several experimental studies examining L_{Max} and L_{Min} observer models have been conducted. As a general summary statement, we have found

support for *both* L_{Max} and L_{Min} mechanisms in different conditions. A portion of the evidence supporting each will be reviewed briefly below. A general observation, though, about the approach to studying these hypothetical observer models: it is difficult to devise any experiment that conclusively rules out one particular strategy. Instead, the approach that is usually taken is one in which the control condition randomizes both target and masker values along the relevant dimension. The comparison condition then holds constant the value of one of the types of stimuli - target or masker - so that the observer knows beforehand which stimulus should receive particular emphasis/preprocessing. Thus, if advantages are found for fixed-target conditions, we infer that the observer adopts or exploits an L_{Max} strategy, maximizing the target properties. Conversely, if the masker value is fixed any performance advantages may be attributed to an L_{Min} process. Although somewhat indirect, this approach yields sensible patterns of results. The fact that the different experiments support *either* L_{Max} or L_{Min} makes it difficult to propose a single model to account for listener performance. However, the notion that human observers have multiple strategies available for use in solving complex listening tasks - and apply them according to circumstance - is neither surprising nor new. The challenge is to find consistencies in the patterns of results implicating broad categories of task demands for which one strategy or the other is optimal.

The first example of support for an L_{Max} model may be found by inspection of Figure 3 above. Using the every-other-word speech identification task, Kidd et al. (2008b) compared performance in a control condition in which the relevant parameters of both target and masker words were randomized across trials with comparison conditions in which either or both target and masker values were fixed across trials. In the Kidd et al. study, correct syntactic structure was one such "linkage variable" (as per Figure 3) as were constant talker voice and apparent spatial location. In all cases, Kidd et al. found significant benefits when the target values were held constant but no corresponding benefits when masker values were held constant. Thus, they concluded that their findings were consistent with an L_{Max} observer strategy with no support found for an L_{Min} strategy.

A second recent study using nonspeech stimuli also found support for an L_{Max} observer model. In this case, Kidd et al. (2008c) measured detection thresholds under certain and uncertain target frequency conditions. In the certain condition, the target tone frequency was held constant ("fixed") across each block of trials while in the uncertain condition the target frequency was chosen randomly on each trial ("random"). Thresholds for two target frequencies were measured; one at a relatively low frequency and one at a relatively high frequency. Three conditions were tested for each: an unmasked control, a notched-filtered Gaussian noise

Fig. 5



masker, and a multitone masker whose frequencies were randomly selected presentation-by-presentation (cf. Neff and Green, 1987; review in Kidd et al., 2008a). The notched-noise and multitone maskers are illustrated schematically in **Figure 5** along with the two target frequencies (red lines inside of "protected regions" where masker energy was excluded).

In the random-frequency condition, the listener had to monitor two frequency regions and make detection judgments about each while in the fixed-target frequency case only one frequency region had to be monitored by the listener. The assumption was that any benefit to detection performance due to holding target frequency constant could be related to an L_{Max} strategy in which the observer emphasized the processing at the known (and therefore attended) frequency region. **Figure 6** illustrates the results. This figure shows thresholds for fixed-frequency and random-frequency targets (upper panel) in quiet, notched-noise and multitone maskers as well as the "costs" (difference between fixed- and random-frequency thresholds; lower panel) for each. The important finding from this study was the much larger costs associated with target frequency uncertainty for the highly informational multitone masker. This finding is still not fully understood and work continues to explain and model the results. Conversely, one can think of the costs of uncertainty as indicating the benefit afforded by an L_{Max} observer strategy.

Fig. 6

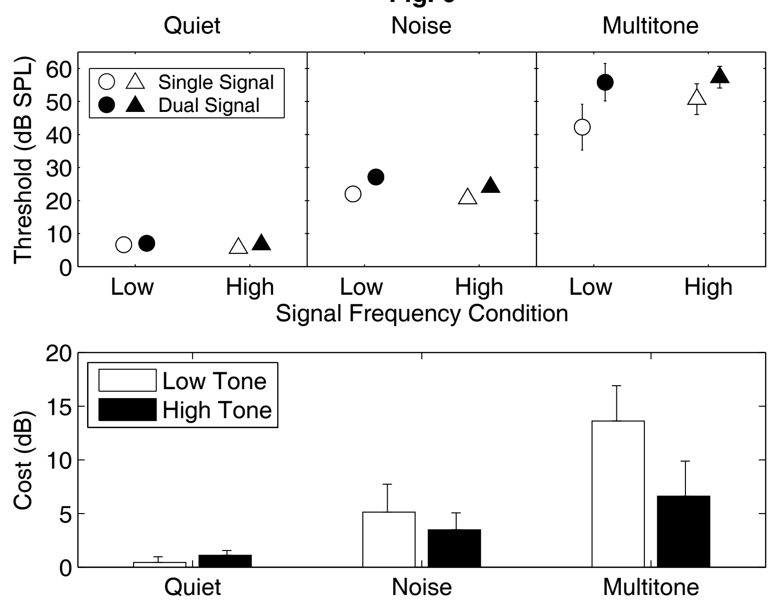
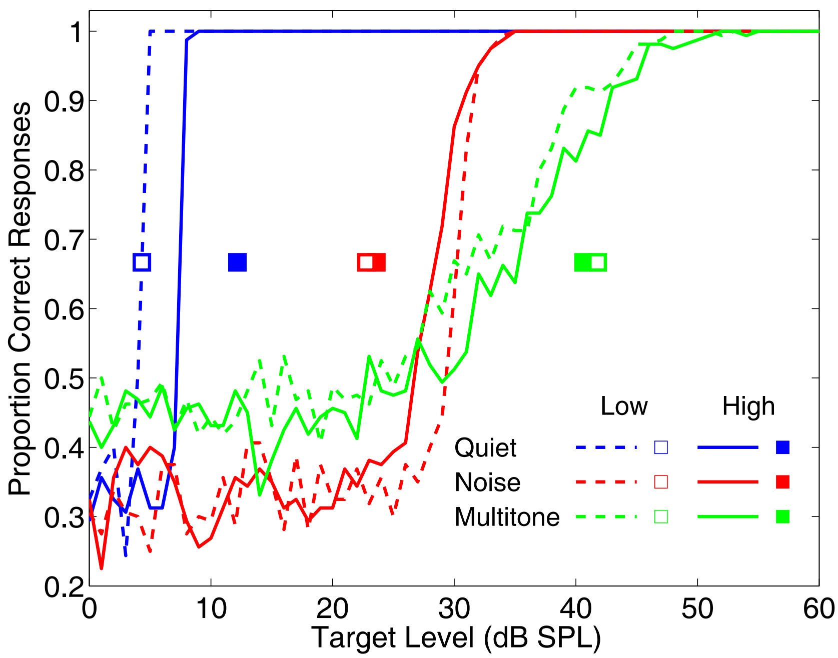


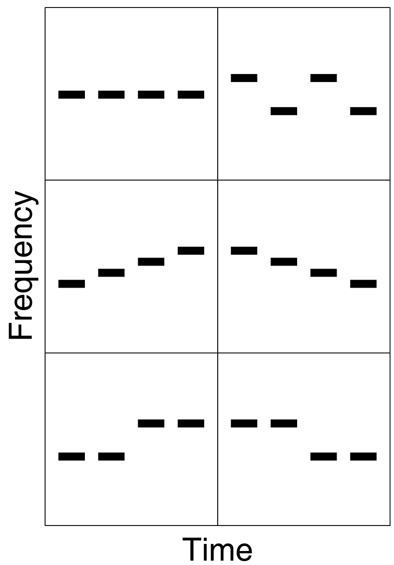
Fig. 7



Recently, we have begun devising a model to account for these effects (Thompson and Kidd, 2011). To date, the initial efforts capture some of the important effects for the differences in masking due to energetic (notched-noise) vs. informational (randomized multitone) maskers, but do not yet successfully predict the costs for the different masker types. However, for the fixed-frequency conditions, the thresholds and slopes of the underlying psychometric functions are predicted to a rough first-order approximation. This is illustrated in **Figure 7**. Shown here are the model predictions plotted as psychometric functions (solid and dashed lines) along with the corresponding group mean data points (squares) from Thompson and Kidd (2011). This model uses the physiologically-inspired preprocessing front end described by Dau et al. (1996) that consists of an outer/middle ear transfer function, gammatone cochlear filter bank, half-wave rectification and low-pass filtering, and adaptation loops. The decision process is based on an ideal detector in which the signal is known exactly and the masker statistics are known and stationary (cf. Green and Swets, 1974). The model captures the difference in slope with masker type (data for slopes not shown, cf., Kidd et al., 1998, 2002). We are currently testing an alternative approach based on Lutfi's CoRE model (Lutfi, 1993; Alexander and Lutfi, 2004) exploring modifications that can capture the added costs associated with the interaction between target and masker frequency uncertainty.

The preceding findings may be construed as providing evidence for the benefit of an L_{Max} observer model. This is because a priori knowledge about the target in a highly uncertain informational masking listening situation improved performance. However, as noted above, we have also found equally convincing evidence in support of an L_{Min} observer model. Kidd et al. (2011) examined contextual effects in the identification of complex, nonspeech sounds. The identification task, which we have used in past studies of energetic and informational masking,

Fig. 8



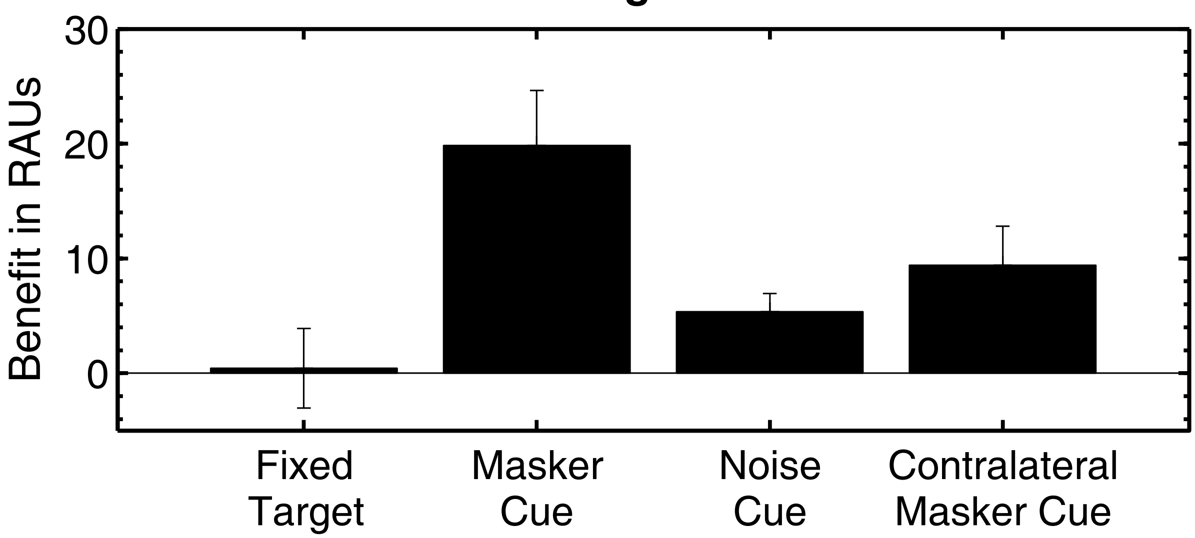
requires subjects to learn six pure-tone sequences that differ in the order of frequencies within a narrow band. A schematic illustration of this set of spectrotemporal patterns is shown in **Figure 8**. Note that these representations are of *relative* frequency. The narrowband patterns (e.g., total range usually about 14% of the nominal "center frequency") are easily identifiable in quiet regardless of the absolute frequency at which they are presented. This makes this stimulus set a good choice for studying mechanisms of masking at suprathreshold levels because masker energy may be overlaid directly on the patterns causing predominantly energetic masking or may be presented remote in frequency to the targets creating informational masking (e.g., Kidd et al., 1998, 2002). They are also well-suited for examining auditory processing of sequences of sounds, especially when varying target and/or masker uncertainty is of interest.

Figure 9 illustrates one trial of an experiment that used the nonspeech pattern identification task to examine sequential interactions among stimuli (Kidd et al. 2011). This figure is a schematic in sound spectrogram form of a target (bold, one of the six patterns shown in Figure 8) masked by a multitone masker randomized in frequency content from trial to trial. Except for one condition discussed below, the target frequency was also chosen randomly on every trial. The task is to identify the target pattern in a 1I6AFC paradigm. In this illustration, the target-plus-masker (right portion of panel) is preceded by a cue/precursor. In the schematic, the precursor shown is an exact copy of the subsequent masker. The contextual effects that Kidd

Fig. 9

et al. were interested in studying tend to emphasize or "enhance" spectral contrasts (cf. the "enhancement effect": Viemeister 1980; Viemiester and Bacon, 1982; Summerfield et al., 1987; Byrne et al., 2011). Various precursors were tested including the exact copy of the masker shown in Figure 9 ("masker cue") as well as that precursor presented to the opposite ear from the target-plus-masker ("contralateral masker cue") and a notched-filtered noise having a notch centered on the target presented ipsilateral to the target-plus-masker ("noise cue"). One condition without a precursor tested the effectiveness of an L_{Max} strategy by holding the target frequency constant across trials ("fixed target"). The improvement in performance from each of these contextual cues, relative to an uncued control condition in which both the target and masker were randomized across trials, is plotted in **Figure 10** in rationalized arcsine units (RAUs). First of all, no significant benefit was found for holding the target frequency constant. In this highly uncertain experiment, this result does not support the actions of an L_{Max} listener strategy. In contrast, the three masker precursors did provide a significant benefit, with the greatest advantage found for the exact masker precursor. That finding supports an effective L_{Min} strategy. However, a portion of the effect - that revealed by the small but significant notched-

Fig. 10



noise benefit - may reflect a component of auditory enhancement that depends solely on differential prior stimulation of masker and target channels. This enhancement effect likely is due to bottom-up inhibitory processes that are not directed by attentional control or a priori knowledge. In contrast, the somewhat larger contralateral exact-masker cue benefit may be complementary and directed by top-down mechanisms. This latter finding may be related to the contralateral contrast enhancement effects in sequences of speech sounds reported by Holt and colleagues (Holt, 2005, 2006a,b; Lotto et al., 2003).

III. Peer Reviewed Articles (2008-present)

Kidd, G. Jr., Richards, V.M, Streeter, T., Mason, C.R. and Haung, R. (2011) "Contextual effects in the identification of nonspeech auditory patterns," J. Soc. Am., 130, 3926-3938

Best, V., Mason, C.R. and Kidd, G. Jr., (2011) "Spatial release from masking as a function of the temporal overlap of competing maskers", J. Acoust. Soc. Am., 129, 1616-1625

Kidd, G. Jr., Mason, C.R., Best, V. and Marrone, N.L. (2010) "Stimulus factors influencing spatial release from speech on speech masking," J. Acoust. Soc. Am., 128, 1965-1978

Best, V., Gallun, F.J., Mason, C.R., Kidd, G. Jr. and Shinn-Cunningham, B.G. (2009) "The impact of noise and hearing loss on the processing of simultaneous sentences," Ear and Hearing, 31, 213-220

Kidd, G. Jr., Streeter, T., Ihlefeld, A. and Mason, C.R. (2009) "The intelligibility of pointillistic speech," J. Acoust. Soc. Am. 126, EL186-201

Best, V., Marrone, N., Mason, C.R., Kidd Jr., G. and Shinn-Cunningham, B.G. (2009) "Effects of Sensorineural Hearing Loss on Visually Guided Attention in a Multitalker Environment," J. Assoc. Res. Oto. 10 142-149

Kidd, G. Jr., Best, V. and Mason, C.R. (2008) "Listening to every other word: Examining the strength of linkage variables in forming streams of speech" J. Acoust. Soc. Am. 124, 3793-3802

Fan, W.L., Streeter, T.M. and Durlach, N.I. (2008) "The effect of spatial uncertainty for masker on masked detection for nonspeech stimuli," J. Acoust. Soc. Am. (L), 124, 36-39

Kidd, G. Jr., Richards, V.M., Mason, C.R., Gallun, F.J. and Huang, R. (2008) "Informational masking increases the costs of monitoring multiple channels," J. Acoust. Soc. Am., 124, EL223-229

Gallun, F.J., Durlach, N.I., Colburn, H.S., Shinn-Cunningham, B.G., Best, V., Mason, C.R. and Kidd, G. Jr. (2008) "The extent to which a position-based explanation accounts for binaural release from informational masking" J. Acoust. Soc. Am., 124, 439-449

Marrone, N.L., Mason, C.R. and Kidd, G. Jr. (2008) "Tuning in the spatial dimension: Evidence from a masked speech identification task" J. Acoust. Soc. Am., 124, 1146-1158

IV. Papers Presented at Scientific Meetings

Kidd, G. Jr., Richards, V., Streeter, T., Mason, C., and Huang, R. (2011) Contextual effects in the identification of nonspeech auditory patterns in random maskers. Presented at the Midwinter Meeting of the Association for Research in Otolarygology, Baltimore, MD.

Thompson, E. and Kidd, G. Jr. (2011) Interactions between target uncertainty, masker uncertainty and hearing loss. Presented at the Midwinter Meeting of the Association for Research in Otolarygology, Baltimore, MD.

Thompson, E. R., Mason, C.R., Kidd, G. Jr. (2011) Temporal effects with multiple-burst multitone masking. J. Acoust. Soc. Am., 129, 2588.

Best, V., Thompson, E.R., Mason, C.R., Kidd, G. Jr. (2011) Interactions between hearing loss and the spectral overlap of competing sounds in masked speech intelligibility. J. Acoust. Soc. Am. 129, 2590.

Best, V., Marrone, N., Mason, C. and Kidd, G. Jr. (2010) Release from speech-on-speech masking under degraded signal conditions. J. Acoust. Soc. Am., 127, 1809.

Iyer, N., Best, V., Brungart, D., Simpson, B., Mason, C. and Kidd, G. (2009) Factors influencing the sequential organization of speech. Presented at the Midwinter Meeting of the Association for Research in Otolarygology, Baltimore, MD.

Best, V., Mason, C. and Kidd, G. (2009) Spatial release from speech-on-speech masking as a function of temporal overlap in listeners with sensorineural hearing loss. Presented at the Midwinter Meeting of the Association for Research in Otolarygology, Baltimore, MD.

Kidd, G. Jr., Best, V., Mason, C.R. (2008) Listening to every other word: Testing the strength of linkage variables in forming streams of speech. Presented at the Midwinter Meeting of the Association for Research in Otolaryngology, Phoenix, AZ.

Kidd, G. Jr., Mason, C.R., Best, V., Marrone, N.L. and Durlach, N.I. (2008) An investigation of the factors responsible for spatial release from masking. Presented at the Midwinter Meeting of the Association for Research in Otolaryngology, Phoenix, AZ.

V. REFERENCES

Akeroyd, M. A. (2004). “The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking,” J. Acoust. Soc. Am. 116,1135–1148

Alexander, J.M. and Lutfi, R.A. (2004) "Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights," J. Acoust. Soc. Am., 116, 2234-2247

Allen, K., Carlile, S. and Alais, D. (2008) “Contributions of talker characteristics and spatial location to auditory streaming,” J. Acoust. Soc. Am. 123, 1562-1570

Arbogast, T.L., Mason, C.R., and Kidd, G. Jr. (2002) "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am., 112, 2086-2098

Best, V., Ozmeral, E., Gallun, F.J., Sen, K. and Shinn-Cunningham, B.G. (2005) “Spatial unmasking of birdsong in human listeners: Energetic and informational factors”, J. Acoust. Soc. Am. 118, 3766-3773

Best, V., Mason, C.R. and Kidd, G. Jr., (2011) "Spatial release from masking as a function of the temporal overlap of competing maskers", J. Acoust. Soc. Am., 129, 1616-1625

Bozzi, P.,& Vicario, G. (1960). “Due fattori di unificazione fra note musicali: la vicinanza temporale e la vicinanza tonale”, Rivista di Psicologia, ann. LIV, n. IV, 235-258

Bregman, A. *Auditory Scene Analysis: The Perceptual Organization of Sounds*, Bradford Press, MIT Press, Cambridge, MA, 1990

Bregman, A.S. "Auditory scene analysis," in A.I. Basbaum, A. Koneko, G.M. Shepherd & G.Westheimer (Eds.) *The Senses: A Comprehensive Reference, Vol. 3, Audition*, P. Dallos & D. Oertel (Volume Eds.) San Diego: Academic Press, 2008, Pp. 861-870

Broadbent, D.E. (1952) "Failures of attention in selective listening," J. Exp. Psychol., 44, 428-433

Brown, C.A. and Bacon, S.P. (2009) "Achieving electric-acoustic benefit with a modulated tone," Ear Hear., 30, 489-493

Brungart, D.S., Simpson, B.D., Darwin, C.J., Arbogast, T.L. and Kidd, G. Jr. (2005) "Across-ear interference from parametrically-degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am., 117, 292-304

Brungart, D.S., Chang, P.S., Simpson, B.D. and Wang, D. (2006). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. 120, 4007-4018

Brungart, D.S., Iyer, N. and Simpson, B.D. (2007) "Selective spatial attention in a dynamic cocktail party task: Evidence for a strategy based on masker minimization," J. Acoust. Soc. Am., 121, 3119

Brungart, D.S., Chang, P.S., Simpson, B.D. and Wang, D. (2009) "Multitalker speech perception with ideal time-frequency segregation: effects of voice characteristics and number of talkers," J. Acoust. Soc. Am., 125, 4006-22

Byrne, A.J., Stellmack, M.A. and Viemeister, N.F. (2011) "The enhancement effect: Evidence for adaptation of inhibition using a binaural centering task," J. Acoust. Soc. Am., 129, 2088-2094

Cherry EC (1953) "Some experiments on the recognition of speech, with one and with two ears", J. Acoust. Soc. Am., 25: 975-979.

Colburn, H. S. (1995). “Computational models of binaural processing,” in *Auditory Computation*, edited by H. Hawkins and T. McMullin (Springer-Verlag, New York), pp. 332–400

Colburn, H. S., and Durlach, N. I. (1978). “Models of binaural interaction,” in *Handbook of Perception: Hearing*, edited by E. Carterette and M. Friedman (Academic Press, New York), Vol. 4, Chap. 11, pp. 467–518

Culling, J.F. and Colburn, H.S. (2000) Binaural sluggishness in the perception of tone sequences and speech in noise. J. Acoust. Soc. Am. 107, 517-527.

Culling, J.F., Edmonds, B.A. and Hodder, K.I. (2006) "Speech perception from monaural and binaural information," J. Acoust. Soc. Am. 119, 559-565

Dau, T., Puschel, D. and Kohlrausch, A. (1996) "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am., 99, 3615-3622

Dorman, M.F., Loizou, P.C. and Rainey, D. (1997) "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am., 102, 2403-2411

Durlach, N. I. (1963). “Equalization and cancellation theory of binaural masking-level differences,” J. Acoust. Soc. Am. 35, 1206–1218

Durlach, N. I. (1972). “Binaural signal detection: Equalization and cancellation theory,” in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic Press, New York), Vol. 2, Chap. 10, pp. 369–462.

Durlach, N. I., and Colburn, H. S. (1978). “Binaural phenomena,” in *Handbook of Perception: Hearing*, edited by E. Carterette and M. Friedman (Academic Press, New York), Vol. 4, Chap. 10, pp. 405–466.

Durlach, N.I., Mason, C.R. , Kidd, G. Jr., Arbogast, T.L., Colburn, H.S. and Shinn-Cunningham, B.G. (2003) "Note on informational masking " J. Acoust. Soc. Am. (L), 113, 2984-2987.

Engle, R.W. (2010) "Role of working-memory capacity in cognitive control," Current Anthropology, 51, Suppl. 1, S17-S26

Gallun, F.J., Mason, C.R. and Kidd, G. Jr. (2005) "Binaural release from informational masking in a speech identification task," J. Acoust. Soc. Am., 118, 1614-1625.

Green, D.M. and Swets, J.A ., *Signal Detection Theory and Psychophysics*, Krieger Publishing Company, 1974 (Orig. pub. Wiley, New York, 1966).

Holt, L.L. (2005) "Temporally nonadjacent nonlinguistic sounds affect speech categorization," Psychol. Sci., 16, 305-312

Holt, L.L. (2006a) "Speech categorization in context: Joint effects of speech and nonspeech precursors," J. Acoust. Soc. Am., 119, 4016-4026

Holt, L.L. (2006b) "The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization," J. Acoust. Soc. Am., 120, 2801-2817

Iyer, N., Best, V., Brungart, D., Simpson, B., Mason, C.R. and Kidd, G. Jr., (2009) "Factors influencing the sequential organization of speech," Presented at the 2009 Midwinter Research Meeting of the Association for Research in Otolaryngology, Baltimore, MD

Kidd G, Jr., Mason C.R., Rohtla, T.L. and Deliwala, P.S. (1998) "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J Acoust Soc Am, 104, 422-431

Kidd G, Jr., Mason C.R. and Arbogast, T.L. (2002) "Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am.,111, 1367-1376.

Kidd, G. Jr., Mason, C.R., Richards, V. M., Gallun, F. J. and Durlach, N.I. "Informational masking," in *Auditory Perception of Sound Sources*, Yost, W. A., Popper, A. N., and Fay, R. R. (eds). (2008a). Springer Science+Business Media, LLC, New York, 143-190

Kidd, G. Jr., Best, V. and Mason, C.R. (2008b) "Listening to every other word: Examining the strength of linkage variables in forming streams of speech" J. Acoust. Soc. Am. 124, 3793-3802

Kidd, G. Jr., Richards, V.M., Mason, C.R., Gallun, F.J. and Huang, R. (2008c) "Informational masking increases the costs of monitoring multiple channels," J. Acoust. Soc. Am., 124, EL223-229

Kidd, G. Jr., Streeter, T., Ihlefeld, A. and Mason, C.R. (2009) “The intelligibility of pointillistic speech,” J. Acoust. Soc. Am. 126, EL186-201

Kidd, G. Jr., Mason, C.R., Best, V. and Marrone, N.L. (2010) “Stimulus factors influencing spatial release from speech on speech masking,” J. Acoust. Soc. Am., 128, 1965-1978

Kidd, G. Jr., Richards, V.M, Streeter, T., Mason, C.R. and Haung, R. (2011) "Contextual effects in the identification of nonspeech auditory patterns," (in revision, J. Acoust. Soc. Am.)

Kong, Y. and Carlyon, R.P. (2007) "Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation," J. Acoust. Soc. Am., 117, 1351-1361

Levitt, H. and Rabiner, L.R. (1967) “Binaural release from masking and gain in intelligibility”, J. Acoust. Soc. Am., 42, 601-508.

Lotto, A.J., Sullivan, S.C. and Holt, L.L. (2003) “Central locus for nonspeech context effects on phonetic identification,” J. Acoust. Soc. Am. (L), 113, 53-56

Lutfi, R.A. (1993) "A model of auditory pattern-analysis based on component-relative-entropy," J. Acoust. Soc. Am., 94, 748-758.

Marrone, Nicole L. *The benefit of spatial separation between multiple talkers: A comparison of aided and unaided listening in reverberant rooms,* Ph.D. Dissertation, Boston University, 2007

Marrone, N.L., Mason, C.R. and Kidd, G. Jr. (2008a) "Tuning in the spatial dimension: Evidence from a masked speech identification task" J. Acoust. Soc. Am., 124, 1146-1158

Marrone, N.L., Mason, C.R. and Kidd, G. Jr. (2008b) ""Effect of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms" J. Acoust. Soc. Am., 124, 3064-3075

Marrone, N.L., Mason, C.R. and Kidd, G. Jr. (2008c) "Evaluating the benefit of hearing aids in solving the cocktail party problem," Trends in Amplification, 12, 300-315

Moore, B.C.J. and Gockel, H. (2002) "Factors influencing sequential stream segregation," Acta Acoust United with Acoustica, 88, 320-332

Neff, D.L. and Green, D.M. (1987) "Masking produced by spectral uncertainty with multicomponent maskers," Percept Psychophys, 41, 409-415

Pollack, I. (2002) Personal communication.

Pollack, I. (1976) "Identification of random auditory waveforms. III. Effect of interference," J. Acoust. Soc. Am., 60, 680-686

Qin, M.K. and Oxenham, A.J. (2006) "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," J. Acoust. Soc. Am. 119, 2417-2426

Rabiner, L.R. (1989) "A tutorial on Hidden Markov Models and selected applications in speech recognition," Proceed. IEEE, 77, 257-286

Shannon, R.V., Zeng, F-G, Wygonski, J., Kamath, V. and Ekelid, M. (1995) "Speech recognition with primarily temporal cues," Science, 270, 303–304

Sterian, A. *Model-based Segmentation of Time-frequency Images for Musical Transcription*, Ph.D. Thesis, Department of Electrical Engineering, University of Michigan, 1999

Sterian, A. and Wakefield, G. H. (1996). "Robust automated music transcription systems," Proc. of Intl. Comp. Music Conf., Hong Kong

Stern, R. M., and Trahiotis, C. (1996). “Models of binaural perception,” in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates, New York), pp. 499–531.

Stickney, G.S., Assman, P.F., Chang, J. and Zhang, F.-G. (2007) “Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences,” J. Acoust. Soc. Am., 122, 1069-107

Summerfield, A.Q., Sidwell, A. and Nelson, A. (1987) "Auditory enhancement of changes in spectral amplitude," J. Acoust. Soc. Am., 81, 700-708

Thompson, E. and Kidd, G. Jr., (2011) "Interactions between target uncertainty, masker uncertainty and hearing loss," Program of the 34th Annual Midwinter Research Meeting of the Association for Research in Otolaryngology

Thompson, E. (2011) Unpublished observations.

van Noorden, L.P.A.S. *Temporal Coherence in the Perception of Tone Sequences*, Ph.D. Dissertation, Eindhoven University of Technology, 1975

Viemeister, N.F. "Adaptation of masking," in *Psychophysical, Physiological and Behavioral Studies in Hearing*, G. van den Brink and F.A. Bilsen (eds), Noordwijkerhout, The Netherlands, 1980, pp.190-198

Viemeister, N. F. and Bacon, S. P. (1982) “Forward masking by enhanced components in harmonic complexes,” J. Acoust. Soc. Am., 71, 1502-1507

Wan, R., Durlach, N.I. and Colburn, H.S. (2010) "Application of an extended equalization-cancellation model to speech intelligibility with spatially distributed maskers," J. Acoust. Soc. Am. 128, 3678-3690

Watson, C.S. (2005) "Some comments on informational masking," Acta Acustica united with Acustica, 91, 502 – 512.

Watson, C. and Kelly, W. (1981) "The role of stimulus uncertainty in the discrimination of auditory patterns," In: Getty DJ, Howard JH editors. *Auditory and visual pattern recognition*, Hillsdale, NJ: Lawrence Erlbaum. p 37–59.

Watson, C.S. (1987) "Uncertainty, informational masking, and the capacity of immediate auditory memory," In: Yost WA, Watson CS editors. *Auditory processing of complex sounds*.: Lawrence Erlbaum Associates, Inc.

Yost, W.A. (1991) "Auditory image perception and analysis: the basis for hearing," Hear. Res., 56, 8-18

Yost, W.A. "Perceiving sound sources," in *Auditory Perception of Sound Sources*, W.A. Yost, A.N. Popper and R.R. Fay (Eds), Springer Handbook of Auditory Research, 11-13, 2008

Zurek, P.M. (1993) "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance, 2nd Edition,* G.A. Studebaker and I. Hochberg (Eds), Allyn and Bacon, Needham Heights, MA.